

10,827,325

日本国特許庁
JAPAN PATENT OFFICE

07.20.2004

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日 2004年 2月 4日
Date of Application:

出願番号 特願2004-027490
Application Number:

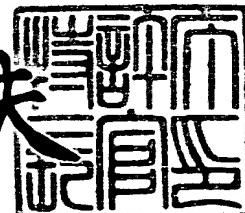
[ST. 10/C] : [JP 2004-027490]

願人 株式会社日立製作所
Applicant(s):

CERTIFIED COPY OF
PRIORITY DOCUMENT

2004年 3月 31日

今井康夫



特許庁長官
Commissioner,
Japan Patent Office

【書類名】 特許願
【整理番号】 K04002791A
【あて先】 特許庁長官殿
【国際特許分類】 G06F 3/06
【発明者】
【住所又は居所】 神奈川県小田原市中里 322 番 2 号 株式会社日立製作所 R A
I D システム事業部内
【氏名】 森田 星輝
【特許出願人】
【識別番号】 000005108
【氏名又は名称】 株式会社 日立製作所
【代理人】
【識別番号】 100075096
【弁理士】
【氏名又は名称】 作田 康夫
【選任した代理人】
【識別番号】 100100310
【弁理士】
【氏名又は名称】 井上 学
【手数料の表示】
【予納台帳番号】 013088
【納付金額】 21,000円
【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1

【書類名】特許請求の範囲**【請求項1】**

ディスクアレイ装置であって、
ディスクアレイ装置筐体と、
前記ディスクアレイ装置筐体内に格納された複数のディスク装置と、
前記ディスクアレイ装置筐体内に格納され、前記ディスク装置へのデータの読み書きを制御するコントローラと、
前記各ディスク装置およびコントローラを接続するケーブルとを有し、
前記ディスク装置は、第1のディスク装置と、前記第1のディスク装置のインターフェースと異なるインターフェースを有する第2のディスク装置を有し、
前記コントローラが、前記第1のディスク装置の故障を判断した場合に、前記第2のディスク装置を用いてスペアリングを行うことを特徴とするディスクアレイ装置。

【請求項2】

請求項1に記載のディスクアレイ装置であって、
前記第1のディスク装置のインターフェースはファイバチャネルインターフェースであり、
前記第2のディスク装置のインターフェースはシリアルインターフェースであることを特徴とするディスクアレイ装置。

【請求項3】

請求項1記載のディスクアレイ装置であって、
前記コントローラは、
前記各ディスク装置の読み書き時に生じるエラー回数に基づいて、該ディスク装置を閉塞するか否かを判断する判断部と、
閉塞すると判断した時に、該閉塞されるディスク装置のスペアとして、前記複数のディスク装置の一部を割り当てるスペアリング処理を制御するスペアリング制御部と、
該閉塞の発生を、所定の通知タイミングで、予め設定された連絡先に通知する異常通知部とを備え、
該異常通知部は、同一種類のディスク装置間で前記スペアリング処理が行われる場合の通知タイミングよりも、異種のディスク装置間で前記スペアリング処理が行われる場合の通知タイミングの方が早くなるよう、前記通知タイミングを設定するディスクアレイ装置。

【請求項4】

請求項3記載のディスクアレイ装置であって、
前記スペアリング制御部は、前記同一種類のディスク装置間でのスペアリング処理を、前記異種のディスク装置間でのスペアリング処理よりも優先して実行するディスクアレイ装置。

【請求項5】

請求項3記載のディスクアレイ装置であって、
前記異常通知部は、更に、閉塞されたディスク装置の台数、および前記スペアリング処理で割り当て可能なディスク装置の台数の少なくとも一方に基づいて、前記通知タイミングを設定するディスクアレイ装置。

【請求項6】

請求項3記載のディスクアレイ装置であって、
前記異常通知部は、前記ディスクアレイ装置について、前記閉塞以外の支障も通知可能であり、前記通知タイミングに至る前に該支障が生じた場合には、前記支障の通知と併せて前記閉塞の通知を行うディスクアレイ装置。

【請求項7】

請求項3記載のディスクアレイ装置であって、
前記スペアリング制御部は、前記異種のディスク装置間でのスペアリング処理時に、前記種類間の特性の差違を補償するようディスク装置の割り当てを制御するディスクアレイ装置。

【請求項8】

請求項7記載のディスクアレイ装置であって、

前記ディスク装置には、ファイバチャネルインタフェースを有するファイバチャネルディスク装置と、シリアルインタフェースを有するシリアルディスク装置の2種類が含まれ

前記スペアリング制御部は、前記ファイバチャネルディスク装置を閉塞させる場合には、前記複数のシリアルディスク装置を並列的に割り当てるディスクアレイ装置。

【請求項9】

請求項7記載のディスクアレイ装置であって、

前記ディスク装置には、ファイバチャネルインタフェースを有するファイバチャネルディスク装置と、シリアルインタフェースを有するシリアルディスク装置の2種類が含まれ

前記スペアリング制御部は、前記シリアルディスク装置を閉塞させる場合には、前記複数のファイバチャネルディスク装置を直列的に割り当てるディスクアレイ装置。

【請求項10】

請求項3記載のディスクアレイ装置であって、

前記ディスク装置には、ファイバチャネルインタフェースを有するファイバチャネルディスク装置と、シリアルインタフェースを有するシリアルディスク装置の2種類が含まれ

前記ケーブルは、ファイバチャネルであり、

前記シリアルディスク装置は、前記シリアルインタフェースを前記ファイバチャネルインターフェースに変換するための変換器を介して前記ファイバチャネルに接続されるディスクアレイ装置。

【請求項11】

請求項10記載のディスクアレイ装置であって、

前記コントローラを複数有し、

各コントローラは前記ファイバチャネルケーブルによって相互に接続されるとともに、個別に各ディスク装置と接続されることで、複数のファイバチャネルループを構成し、前記各シリアルディスク装置と前記複数のファイバチャネルケーブルとの間に介在し、前記シリアルディスク装置の接続先を前記複数のファイバチャネルループ間で切り換える切換器とを有するディスクアレイ装置。

【書類名】明細書

【発明の名称】ディスクアレイ装置における異常通知制御

【技術分野】

【0001】

本発明は、異種類のディスク装置が混在しているディスクアレイ装置に関する。例えば、ディスク装置の一部に異常が生じた場合に、異種類のディスクを用いてスペアリングを行うことが可能なディスクアレイ装置及びそのスペアリング方法に関する。

【背景技術】

【0002】

ディスクアレイ装置には、多数のディスク装置が格納されている。これらのディスク装置の一部に障害が生じると、ディスクアレイ装置の動作は保証されない恐れがある。ディスクアレイ装置の耐障害性を向上させるための一つの手段として、スペアリングが挙げられる。スペアリングとは、予めディスクアレイ装置内にスペアのディスク装置を用意しておき、障害が検出されると、速やかにそのディスク装置を閉塞して、スペアのディスク装置を使用する技術を言う。スペアリングが行われると、管理者に異常通知が発信され、保守作業が行われる。このようにしてスペアのディスク装置を使用しつつ、障害が生じたディスク装置を正常なディスク装置に交換することで、ディスクアレイ装置の運用を停止することなく、保守することが可能となる。

【0003】

特許文献1では、ディスク装置のアクセスエラー回数が所定値を超えた時点で、障害が発生する前に予防的にそのディスク装置を閉塞し、スペアのディスク装置に置換する技術が開示されている。特許文献2は、障害発生時には、閉塞されたディスク装置のデータを、複数のスペアのディスク装置に分割して記憶する技術が開示されている。

【0004】

【特許文献1】特開平5-100801号公報

【0005】

【特許文献2】特開2002-297322号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

ディスク装置には、ファイバチャネルインターフェースを有するファイバチャネルディスク装置（以下、「F Cディスク装置」と称する）、シリアルインターフェースを有するシリアルディスク装置（以下、「S A T Aディスク装置」と称する）など、特性の異なる種々の種類が存在する。ディスクアレイ装置において、複数種類のディスク装置を混在して用いれば、各ディスク装置の特徴を活かすとともに、短所を補償し合うことが可能となる。かかるディスクアレイ装置において、スペアリングを行うためには、それぞれの種類ごとにスペアのディスク装置を用意しておくことが望ましい。

【0007】

しかし、ディスクアレイ装置に格納可能なディスク装置の数には限界がある。従って、それぞれの種類ごとにスペアのディスク装置を用意すれば、一つ一つの種類について使用可能なスペアの数を十分に確保することができないという課題を招く。スペアの数が不十分な状態では、わずかな台数のディスク装置に障害が生じた時点で、スペアのディスク装置の残台数が少なくなるため、頻繁に保守を行う必要が生じ、保守に要する負荷が大きくなるという弊害を招く。本発明は、こうした背景に鑑み、異種のディスク装置を混在させたディスクアレイ装置において、保守負荷の極端な増大を招くことなく、スペアリングを可能とすることを目的とする。

【課題を解決するための手段】

【0008】

本発明は、ディスクアレイ装置筐体内に、複数のディスク装置と、ディスク装置へのデータの読み書きを制御するコントローラとが格納され、相互にケーブルで接続されたディ

スクアレイ装置を対象とする。このディスク装置には、特性の異なる複数種類のディスク装置が混在している。かかるディスクアレイ装置において、本発明では、コントローラが、各ディスク装置の読み書き時に生じるエラー回数に基づいて、ディスク装置を閉塞するか否かを判断する。そして、ディスク装置を閉塞すると判断した時には、閉塞されるディスク装置のスペアとして、複数のディスク装置の一部を割り当てるスペアリング処理を実行する。スペアリングに使用されるディスク装置は、閉塞されるディスク装置と同一種類であっても、異なる種類であってもよい。

【0009】

例えば、本発明に係るディスクアレイ装置は、ディスクアレイ装置筐体と、前記ディスクアレイ装置筐体内に格納された複数のディスク装置と、前記ディスクアレイ装置筐体内に格納され、前記ディスク装置へのデータの読み書きを制御するコントローラと、前記各ディスク装置およびコントローラを接続するケーブルとを有し、前記ディスク装置は、第1のディスク装置と、前記第1のディスク装置のインターフェースと異なるインターフェースを有する第2のディスク装置を有し、前記コントローラが、前記第1のディスク装置の故障を判断した場合に、前記第2のディスク装置を用いてスペアリングを行うことを特徴とするディスクアレイ装置である。

【0010】

閉塞に伴い、コントローラは、閉塞の発生を、所定の通知タイミングで、予め設定された連絡先に通知する。本発明では、通知タイミングは、スペアリングが、同一種類のディスク装置間で行われる場合よりも、異種のディスク装置間で行われる場合の方が早くなるよう設定される。一例として、スペアリングが、異種のディスク装置間で行われた場合には、同時に異常を通知し、同種のディスク装置間で行われた場合には、一定期間、通知を保留するようにしてもよい。

【0011】

本発明によれば、異種のディスク装置間でのスペアリングを許容することにより、スペアとして使用可能なディスク装置の数を十分に確保することができ、保守の間隔が短くなることを抑制できる。但し、異種のディスク装置間でスペアリングが行われた場合、ディスク装置の特性の差違により、十分な性能が確保されない恐れがある。本発明では、このような影響を考慮し、異種のディスク装置間でのスペアリング時には通知タイミングを早めることにより、ディスクアレイ装置の性能低下を抑制することができる。

【0012】

本発明においては、同一種類のディスク装置間でのスペアリング処理を、異種のディスク装置間でのスペアリング処理よりも優先して実行することが好ましい。こうすることで、スペアリング処理に伴うディスクアレイ装置の性能低下を抑制することができる。

【0013】

本発明において、通知タイミングは、更に、閉塞されたディスク装置の台数（以下、「閉塞台数」と称する）、およびスペアのディスク装置の残り台数（以下、「スペア台数」と称する）の少なくとも一方に基づいて、設定してもよい。例えば、閉塞台数が所定数以上となった場合や、スペア台数が所定数以下となった時点で、異常を通知するようにしてもよい。こうすることで、通知遅れによって、ディスクアレイ装置の運用が停止せざるを得ない状態を招く可能性を抑制できる。

【0014】

本発明では、ディスクアレイ装置について、閉塞以外の支障を通知可能としてもよい。かかる場合、通知タイミングに至る前に、この支障が生じた場合には、支障の通知と併せて閉塞の通知を行うようにしてもよい。こうすれば、種々の支障に関する保守を同時期に行うことが可能となり、保守に要する負荷を軽減することができる。

【0015】

本発明において、異種のディスク装置間でのスペアリングが行われる場合には、種類間の特性の差違を補償するようディスク装置の割り当てを制御してもよい。例えば、F Cディスク装置とS A T Aディスク装置間のスペアリングの場合、F Cディスク装置を閉塞さ

せる時は複数のSATAディスク装置を並列的に割り当てる方法を探ってもよい。並列的とは、複数のディスク装置に並行してアクセスが行われる様子を意味する。一般にSATAディスク装置はFCディスク装置よりもアクセス速度が低い。このように並列的に割り当てるにより、アクセス速度の低下を抑制することができる。

【0016】

逆に、シリアルディスク装置を閉塞させる時は複数のファイバチャネルディスク装置を直列的に割り当てる方法を探ってもよい。一般にFCディスク装置はSATAディスク装置よりも低容量である。このように直列的に割り当てるにより、容量の低下を抑制することができる。

【0017】

本発明は、種々のディスクアレイ装置、例えば、FCディスク装置と、SATAディスク装置が混在しているディスクアレイ装置に適用可能である。この構成では、SATAディスク装置に対応づけて、シリアルインターフェースをファイバチャネルインターフェースに変換するための変換器を備えることが好ましい。こうすることで、各ディスク装置のインターフェースをファイバチャネルに統一することができる。

【0018】

更に、ディスクアレイ装置の耐障害性を向上させるためデュアルパス化してもよい。即ち、コントローラを複数有し、各コントローラをファイバチャネルケーブルによって相互に接続するとともに、個別に各ディスク装置と接続することで複数のファイバチャネルループを構成してもよい。SATAディスク装置については、その接続先を複数のファイバチャネルループ間で切り換える切換器を設けることで、デュアルパス化することができる。

【0019】

本発明は、ディスクアレイ装置としてのみならず、ディスクアレイ装置における異常通知の制御方法として構成することもできる。例えば、ディスクアレイ装置における異常発生時の通知を制御する異常通知制御方法であって、ディスクアレイ装置筐体と、前記ディスクアレイ装置筐体内に格納された複数のディスク装置と、前記ディスクアレイ装置筐体内に格納され、前記ディスク装置へのデータの読み書きを制御するコントローラとを有し、前記ディスク装置には、特性の異なる複数種類のディスク装置が混在しており、前記コントローラが実行する工程として、前記各ディスク装置の読み書き時に生じるエラーを評価し、該ディスク装置を閉塞するか否かを判断する判断工程と、閉塞すると判断した時に、該閉塞されるディスク装置のスペアとして、前記複数のディスク装置の一部を割り当てるスペアリング処理を制御するスペアリング制御工程と、該閉塞の発生を、所定の通知タイミングで、予め設定された連絡先に通知する異常通知工程とを備え、該異常通知工程は、同一種類のディスク装置間で前記スペアリング処理が行われる場合の通知タイミングよりも、異種のディスク装置間で前記スペアリング処理が行われる場合の通知タイミングの方が早くなるよう、前記通知タイミングが設定されている異常通知制御方法としてもよい。

【0020】

また、本発明は、かかる制御を実現するためのコンピュータプログラム、およびこのコンピュータプログラムを記録したコンピュータ読み取り可能な記録媒体として構成してもよい。記録媒体としては、フレキシブルディスクやCD-ROM、光磁気ディスク、ICカード、ROMカートリッジ、パンチカード、バーコードなどの符号が印刷された印刷物、コンピュータの内部記憶装置（RAMやROMなどのメモリ）および外部記憶装置等、コンピュータが読み取り可能な種々の媒体を利用できる。

【発明を実施するための最良の形態】

【0021】

本発明の実施例について以下の順序で説明する。

【0022】

A. システム構成：

- B. ディスク種別管理処理：
- C. スペアリング処理：
 - C 1. 障害管理テーブル：
 - C 2. スペアリング処理：
 - C 3. 障害通知処理：

A. システム構成：

図1は実施例としての情報処理システムの概略構成を示す説明図である。情報処理システムは、ストレージ装置1000と、ホストコンピュータHCとは、SAN (Storage Area Network) で接続される。各コンピュータHCは、ストレージ装置1000にアクセスして、種々の情報処理を実現することができる。ローカルエリアネットワークLAN (Local Area Network) には、管理装置10が接続されている。管理装置10は、ネットワーク通信機能を有する汎用のパーソナルコンピュータなどを利用することができ、管理ツール11、即ちストレージ装置1000の動作設定をしたり、ストレージ装置1000の動作状態を監視したりするためのアプリケーションプログラムがインストールされている。

【0023】

ストレージ装置1000の内部には、ストレージ装置筐体の内部に複数のディスク装置筐体200、コントローラ筐体300が格納されている。ディスク装置筐体200は、後述する通り、内部に多数のディスク装置（以下、「HDD」と呼ぶこともある）を格納している。ディスク装置は、パーソナルコンピュータなどで採用されている3.5インチの汎用的なものを利用可能である。コントローラ筐体300は、ディスク装置へのデータの読み書きを制御するためのコントローラを格納している。コントローラ筐体300は、ストレージエリアネットワークSANを介してホストコンピュータHCとデータ授受を行い、ローカルエリアネットワークLANを介して管理装置10とデータ授受を行うこともできる。コントローラ筐体300と、各ディスク装置筐体200は、背面側において、ファイバチャネル用のケーブル（以下、「ENCケーブル」と呼ぶ）で相互に接続されている。

【0024】

図示を省略したが、ストレージ装置筐体には、この他、AC/DC電源、冷却ファンユニット、バッテリーユニットが設けられている。バッテリーユニットは二次電池を内蔵しており、停電時に電力を供給するバックアップ電源として機能する。

【0025】

図2はディスク装置筐体200の斜視図である。前面には、ルーバ210が取り付けられており、その内部には、複数のディスク装置220が配列されている。各ディスク装置220は、前面に引き出すことで着脱、交換可能である。図の上方には、背面側の接続パネルの様子を示した。本実施例では、ディスク装置220は、二つのENCユニット202に分けて格納されている。各ENCユニットには、ENCケーブルのIN側コネクタ203、OUT側コネクタ205がそれぞれ2つずつ設けられている。2つのENCユニット202が格納される結果、ディスク装置筐体200には、合計4つのIN側コネクタ203、OUT側コネクタ205、即ち4本のバス（以下、「FC-ALループ」とも称する）に対応したコネクタが設けられることになる。各コネクタには、上方にLED204が設けられている。但し、図の煩雑化を回避するため、LAN204の符号は、コネクタ203[1]についてのみ付した。ENCユニット202には、LANケーブルを接続するためのLAN用コネクタ206、通信状態を示すためのLED207を設けても良い。

【0026】

図3はディスク装置筐体200の内部構造を模式的に示す説明図である。本実施例では、2種類のインターフェースを有するディスク装置220を利用可能とした。一つは、ファイバチャネル用のインターフェースを有するディスク装置220F（以下、「FCディスク装置」と称する）であり、もう一つは、シリアルインターフェースを有するディスク装置220S（以下、「SATAディスク装置」と称する）である。異なるインターフェースを併用可能とするための回路構成については後述する。以下、単に「ディスク装置220」と

いう場合には、インターフェースの種類を問わない総称を意味し、インターフェースごとに区別する場合には、FCディスク装置220F、SATAディスク装置220Sを用いるものとする。

【0027】

上記2種類のディスク装置は、次の特徴を有している。FCディスク装置220Fは、デュアルポート化されており、2つのバスからの読み書きが可能である。また、SCSI 3 (Small Computer System Interface 3) 規格に規定されるSES (SCSI Enclosure Service) やESI (Enclosure Service I/F) の機能を備えている。SATAディスク装置220Sは、本実施例では、シングルポートであり、SESやESIの機能を有していないものとする。但し、これらの機能を有するSATAディスク装置220Sの適用を排除するものではない。

【0028】

図の下方に、各ディスク装置220F、220Sの側面図を示した。それぞれ、ディスク装置筐体200への着脱に利用されるハンドル222F、222Sと、コネクタ221F、221Sが設けられている。コネクタ221F、221Sは、上下方向位置をずらして設けられている。

【0029】

図の中央に示す通り、ディスク装置筐体200の背面には、ディスク装置220を装着するためのコネクタ231F、231Sが配列されたバックボード230が取り付けられている。コネクタ231FはFCディスク装置220F用であり、コネクタ221SはSATAディスク装置220S用である。コネクタ231F、231Sは、上下一組にして、ディスク装置220の装着位置に対応して左右方向に配列されている。各ディスク装置220F、220Sをディスク装置筐体200の前面から引き出し式に挿入すると、ディスク装置のコネクタ221F、221Sは、その種類に応じて、バックボード230のコネクタ231F、231Sのいずれか一方に装着される。ディスク装置220の種類によって装着されるコネクタを変えることで、後述する通り、インターフェースの差違を補償するための回路の使い分けが実現される。また、コネクタの相違は、各ディスク装置220の種類の判別にも利用することができる。更に、格納されているディスク装置の種類が外部から認識可能としてもよい。例えば、格納されているディスク装置または格納すべきディスク装置の種類に応じて、表示ランプの色などを変えるようにしてもよい。

【0030】

コネクタに接続されると、各ディスク装置220は、ディスク装置筐体200の4本のバスPath0～Path3にそれぞれ接続される。本実施例では、各ディスク装置220は、Path0、3に接続されるものと、Path1、2に接続されるものを交互に配置した構成とした。こうすることで、各ディスク装置220については、4本のバスのうち、2本を介してアクセス可能なデュアルバス構成が実現される。図3に示した構成は、一例に過ぎず、ディスク装置筐体200内部のバス数、ディスク装置220との対応関係は、種々の態様を探ることが可能である。

【0031】

図4はストレージ装置1000の内部構造を模式的に示す説明図である。コントローラ筐体300に内蔵されるコントローラ310の内部構造、およびディスク装置筐体200の内部構造を模式的に示した。コントローラ310は、内部にCPU312、RAMやROMなどのメモリ等を備えている。コントローラ310は、ホストコンピュータHCとの通信インターフェースであるホストI/F311、ディスク装置筐体200との通信インターフェースであるドライブI/F315を有している。ホストI/F311は、ファイバチャネル規格に準拠した通信機能を提供する。ドライブI/F315は、SCSI規格やファイバチャネル規格の通信機能を提供する。これらのインターフェースは、複数ポート設けられていても良い。

【0032】

メモリとしては、ディスク装置220への書き込みデータや読み出しだが記憶され

るキャッシュメモリ313、および制御用の種々のソフトウェアを記憶するためのF L A S H メモリ314（共有メモリと呼ぶこともある）などが含まれる。コントローラ310には、A C / D C 電源の状態監視、ディスク装置220の状態監視、表示パネル上の表示デバイスの制御、筐体各部の温度監視などを行う回路が実装されているが、これらについては図示を省略した。

【0033】

本実施例では、2つのコントローラ310[0]、310[1]によって、先に図3に示した4つのループPath0～Path3が形成される。図4中では、図示の煩雑化を回避するため、これらのうちPath0、3またはPath1、2の組み合わせに相当する2つのループを例示した。各コントローラ310[0]、310[1]は、破線で示すようにパスの切換が可能である。従って、例えば、コントローラ310[0]は、図中の矢印a、bに示すように2つのループのいずれを経由しても各ディスク装置220にアクセスすることが可能である。コントローラ310[1]についても同様である。

【0034】

ディスク装置筐体200には、先に説明した通り、複数のディスク装置220が接続される。F Cディスク装置220Fは、P B C (Port Bypass Circuit) 251、252を介して、2つのF C - A L ループにそれぞれ接続される。

【0035】

一方、S A T A ディスク装置220Sは、D P A (Dual Port Apparatus) 232、インターフェース接続装置（例えばS A T A マスター装置、S A T A Master Deviceなど）233、234およびP B C 251、252を介して、2つのF C - A L ループにそれぞれ接続される。D P A 232は、単一ポートのS A T A ディスク装置220Sを、デュアルポート化するための回路である。D P A 232を用いることにより、S A T A ディスク装置220Sは、F Cディスク装置220Fと同様、いずれのF C - A L ループからのアクセスも受け入れ可能となる。

【0036】

インターフェース接続装置233、234は、S A T A ディスク装置220Sのシリアルインターフェースと、ファイバチャネルインターフェースとの変換を行うための回路である。この変換には、例えば、S A T A ディスク装置220Sにアクセスするために用いられるプロトコルおよびコマンドと、ファイバチャネルで用いられるS C S I プロトコルおよびコマンドとの変換が含まれる。

【0037】

先に説明した通り、F Cディスク装置220Fは、S E S 機能を備えているのに対し、S A T A ディスク装置220Sは、この機能を備えていない。ディスク装置筐体200には、この差違を補償するため、筐体管理部241、242が設けられている。筐体管理部241、242は、内部にC P U 、メモリ、キャッシュメモリなどを備えたマイクロコンピュータであり、ディスク装置筐体200内部の各ディスク装置220からディスク種別、アドレス、動作状態その他の管理情報を収集する。筐体管理部241、242はP B C 251、252を介して2つのF C - A L ループに接続されており、収集した情報を、コントローラ310からのS E S コマンドに応じて、コントローラ310に提供する。本実施例では、コントローラ310がディスク装置220の種別に関わらず統一的な方法で管理情報を取得可能とするため、筐体管理部241、242は、S A T A ディスク装置220SのみならずF Cディスク装置220Fについても管理情報の収集を行うものとした。

【0038】

P B C 251は、F C - A L ループに接続される3つの装置、F Cディスク装置220F、インターフェース接続装置233、および筐体管理部241の間で、F C - A L ループの切り換えを行う。即ち、P B C 251は、コントローラ310からのコマンドに応じて、F Cディスク装置220F、インターフェース接続装置233、および筐体管理部241の一つを選択してF C - A L ループに接続し、他の2つを切り離す。同様にして、P B C 252は、F C - A L ループに接続される3つの装置、F Cディスク装置220F、

インターフェース接続装置234、および筐体管理部242の間で、FC-ALループの切り換えを行う。

【0039】

以上で説明した構造により、本実施例のストレージ装置1000は、次の特徴を有する。第1に、インターフェース接続装置233、234の機能により、各ディスク装置筐体200の内部に、FCディスク装置220F、SATAディスク装置220Sという2種類のディスク装置を混在して格納することができる。第2に、DPA232の機能により、SATAディスク装置220Sについてもデュアルポート化が実現される。第3に、筐体管理部241、242の機能により、コントローラ310は、SATAディスク装置220Sについても管理情報の収集が容易となる。これらの特徴は、図1～4で説明した構成に基づくものであり、本実施例に必須という訳ではない。本実施例は、上述のストレージ装置1000以外にも、上記特徴の一部を有しない構造も含め、種々の構造からなるストレージ装置を適用可能である。

B. ディスク種別管理処理：

図5はディスク種別管理処理のフローチャートである。コントローラ310が、各ディスク装置220の種別、即ちFCディスク装置220FであるかSATAディスク装置220Sであるかを把握、管理するための処理である。左側に、コントローラ310が実行する処理を示し、右側に筐体管理部241、242が実行する処理を示した。

【0040】

この処理が開始されると、コントローラ310は、ディスク種別の確認指示を入力する（ステップS10）。確認指示は、例えば、ユーザがコントローラ310の操作、管理装置10からのコマンドによって明示的に行うものとしてもよいし、ストレージ装置1000の起動を確認指示とみなすようにしてもよい。

【0041】

コントローラ310は、確認指示に応じて、ディスク装置筐体200ごとに、そこに格納されているディスク装置220の種別を、筐体管理部241、242に問い合わせる。筐体管理部241、242は、この問い合わせを入力すると（ステップS20）、各ディスク装置220が接続されているコネクタに基づいて、種別を確認する（ステップS22）。つまり、ディスク装置220が、先に図3で示したコネクタ231Fに接続されている場合には「FCディスク装置」であり、コネクタ231Sに接続されている場合には「SATAディスク装置」であると認識する。筐体管理部241、242は、こうして得られた確認結果を、コントローラ310に通知する（ステップS24）。

【0042】

上述の処理は、筐体管理部241、242のうち、コントローラ310からの問い合わせを受けたいずれか一方のみが行えればよい。また、筐体管理部241、242は、ディスク装置の種別を予め確認・保存しておき、この結果を問い合わせに応じてコントローラ310に通知するようにしてもよい。

【0043】

コントローラ310は、筐体管理部241、242からの通知を受けると、その結果をディスク種別管理テーブルに格納する（ステップS14）。ディスク種別管理テーブルは、各ディスク装置220の種別を管理するために、コントローラ310のFLASHメモリに格納されるテーブルである。図中にディスク種別管理テーブルの内容を例示した。ディスク装置220は、ディスク装置筐体200の番号、ENCユニット202の番号、各ポートに固有のアドレスの組み合わせで特定される。例えば、図中の最上段のレコードは、ディスク装置筐体「#00」番のENCユニット「0」番に格納された、アドレス「#00」のディスク装置220が、「FCディスク装置」であることを意味している。

【0044】

コントローラ310は、全ディスク装置筐体について、以上の処理を繰り返し実行することにより（ステップS18）、各ディスク装置220の種別を確認することができる。以上で説明した実施例のストレージ装置1000によれば、FCディスク装置220Fと

S A T Aディスク装置220Sが、各ディスク装置筐体200の内部に混在して格納されている場合でも、コントローラ310は、その種別を容易に確認、管理することができる。従って、コントローラ310は、F Cディスク装置220F、S A T Aディスク装置220Sの特性を活かし、データの読み書きの制御等を行うことが可能となる。

D. スペアリング処理：

上述した種々の手法で確認されたディスク装置の種別は、ストレージ装置1000の運用管理に活用される。ディスク装置の種別の管理情報を活用して行う処理の一つとしてスペアリングが挙げられる。スペアリングとは、各ディスク装置へのアクセス時に生じるエラーを監視し、故障の前兆が現れているディスク装置については、アクセスが不能となる前に、予め用意されたスペアのディスク装置に切り換える処理である。スペアリングが行われると、コントローラ310は、ディスク装置の保守を促すため、所定のタイミングで管理装置10に、障害通知を行う。

【0045】

スペアリングを行うため、ストレージ装置1000に格納されたディスク装置は、平時のR A I D制御で使用するものと、平時は使用せずスペアとして使用するものに予め区分されている。R A I D用、スペア用の区分は、コントローラ310のF L A S Hメモリ内に保持される「障害管理テーブル」に記憶されている。障害管理テーブルは、各ディスク装置のエラー回数やスペアリングが行われているか否かの区別なども併せて管理する。

D 1. 障害管理テーブル：

図6は障害管理テーブルの構成例を示す説明図である。このテーブルは、各ディスク装置（H D D）に対して、スペアリングに関する種々の情報を記録する。図の上方に示す通り、ディスク装置は、各ディスク装置筐体（D I S K # 0 0 ~ # m）内に複数格納されているから、障害管理テーブルでは、各ディスク装置を、2次元配列（筐体番号、筐体内でのシリアル番号）で表すものとする。図示する通り、ディスク装置筐体D I S K # 0 0 に格納される各ディスク装置は、（0, 0）～（0, n）と表されることになる。

【0046】

障害管理テーブルに記録される情報について説明する。「I / F」とは、各ディスク装置のインターフェースの種別、即ち、各ディスク装置がF Cディスク装置か、S A T Aディスク装置かを示している。「障害回数」は、アクセス時のエラー回数である。この数が50回を超えると、スペアリングの対象と判断される。「50回」という回数は一例に過ぎず、この判断には種々の設定が可能である。

【0047】

「状態」とは、ディスク装置の稼働状態を、「正常」、「閉塞」、「疑似閉塞」の3区分で表している。「閉塞」とはスペアリングで他のディスク装置に切り換えられることによって、使用されなくなった状態を意味する。「疑似閉塞」も同様に、スペアリングの結果、使用されなくなった状態を意味する。但し、「閉塞」状態では障害通知が即時に行われ、「疑似閉塞」では障害通知が保留される点で相違する。本実施例では、インターフェースが同種のディスク装置間でスペアリングが行われた場合は「疑似閉塞」とし、異種のディスク装置間でスペアリングが行われた場合は「閉塞」とした。

【0048】

「スペアリング」は、異常と判断されたディスク装置に対するスペアリングの結果を示している。「完了」は正常にスペアリングが完了したことを意味し、「不可」はスペアのディスク装置が存在しないなどの理由によりスペアリングが行えないことを表している。

【0049】

「スペア」欄において、「Y E S」はディスク装置がスペアとして使用可能であることを示し、「-」はスペアでないこと、即ちR A I Dで通常に使用されることを表している。「スペア使用」が「ON」となっているディスク装置は、スペアリングで使用されていることを表している。「元H D D」は、異常が発見され、スペアリングで割り当てられた元のディスク装置を表している。

【0050】

図の例では、ディスク装置(0, 2)は、障害回数が50回に至ったため、スペアリングが行われ、ディスク装置(0, 5)に切り換えられている。ディスク装置(0, 2)、(0, 5)は、いずれもFCディスク装置であるため、ディスク装置(0, 2)の状態は「疑似閉塞」となる。一方、ディスク装置(m, n-1)は、障害回数が50回に至ったため、スペアリングが行われ、2つのディスク装置(m, n-2)、(m, n)に切り換えられている。2つを使用する理由については、後述する。このスペアリングは、異種間で行われているため、ディスク装置(m, n-1)の状態は「閉塞」となる。また、ディスク装置(0, 4)は、障害回数が100回に至っているが、スペアが存在しないため、スペアリングは「不可」とされている。

【0051】

このようにコントローラ310は、障害管理テーブルを用いて、各ディスク装置の稼働状況を監視し、スペアリングを実行する。以下、コントローラ310が実行する処理内容についてフローチャートで説明する。

D2. スペアリング処理：

図7はスペアリング処理のフローチャートである。この処理は、コントローラ310によってストレージ装置1000の運転中、繰り返し実行される。

【0052】

この処理では、コントローラ310は、各ディスク装置220の閉塞要因、即ちアクセス時のエラー回数を監視する（ステップS40）。このエラー回数が所定数、例えば50回を超える場合に、そのディスク装置220には故障の前兆が現れており「閉塞要因有り」と判断される。閉塞要因の監視は、ディスク装置ごとに行われる。

【0053】

コントローラ310は、いずれかのディスク装置について閉塞要因を見いだした場合、スペアリングが必要と判断し（ステップS42）、スペアとして使用可能なディスク装置が存在するか否かを判断する（ステップS44）。この判断は、先に説明した障害管理テーブルを参照することで行うことができる。複数のディスク装置からなるRAIDグループは同じ種類のインターフェースを有するディスク装置で構成することが望ましい。従って、ディスク装置に故障が生じてスペアリングの必要がある場合には、その故障ディスクが所属するRAIDグループ（ECCグループと呼ぶこともある）のディスク装置のインターフェースを確認することが好ましい。この確認結果および残存しているスペアの種類によって、スペアの有無は、以下の3ケースに別れる。

【0054】

ケース1：閉塞要因が見いだされたディスク装置と同一種類のスペアが有る場合；

ケース2：同一種類のスペアは存在しないが、異種のスペアは存在する場合；

ケース3：スペアが一切存在しない場合

上記分類に従えば、異種のディスク装置へのスペアリングが許容されるが、同種のディスク装置へのスペアリングが優先的に行われることになる。コントローラ310は、ケース1に該当する場合は、同種のスペアのいずれかを選択してスペアリングを行うとともに（ステップS46）、障害管理テーブルの内容を更新する（ステップS48）。この場合は、障害が見いだされたディスク装置は、「疑似閉塞」とされる。

【0055】

ケース2に該当する場合には、異種スペアリング処理として、異種のスペアの中のいずれかを選択してスペアリングを行う（ステップS50）。異種スペアリング処理については、FCディスク装置からSATAディスク装置への切り換え時と、その逆で処理内容が異なるため、処理の詳細については後述する。

【0056】

ケース3に該当する場合には、スペアリングは行わずに、障害管理テーブルを更新する（ステップS48）。ケース3の場合は、障害が見いだされたディスク装置は、スペアリング「不可」とされる。コントローラ310は、以上の処理が完了すると、その結果に応じて、障害通知処理、即ち管理装置10に障害通知を行う処理を実行して（ステップS6

0)、スペアリング処理を完了する。

【0057】

図8は異種スペアリング処理のフローチャートである。図7のステップS50に対応する処理であり、FCディスク装置とSATAディスク装置の間でスペアリングを実行するための処理である。コントローラ310は、この処理が開始されると、障害が生じたディスクの種別を判断する（ステップS52）。なお、異なるインターフェースを有するディスク装置でスペアリングを行わないように、予めユーザ、保守員が障害管理テーブル等に設定をしてもよい。かかる設定がなされている場合には、同種類のスペアディスクが存在しない場合は異種スペアリングは行われない。

【0058】

FCディスク装置で障害が生じている場合には（ステップS52）、コントローラ310は、複数のSATAディスク装置を並列に割り当てることでスペアリングを実行する（ステップS54）。図中に、このスペアリングの様子を模式的に示した。FCディスク装置がRAIDを構成し、SATAディスク装置がスペアとして残っているとする。この状態で、FCディスク装置のひとつに障害が生じた場合、コントローラは2つのSATAディスク装置を並列的に割り当てる。並列に割り当てるとは、これらのSATAディスク装置へのアクセスがほぼ並行して行われるように、データを両者に分散させて格納することを意味する。FCディスク装置に対して3台以上のSATAディスク装置を割り当ててもよい。

【0059】

一般に、アクセス速度は、SATAディスク装置の方がFCディスク装置よりも遅い。従って、一台のFCディスク装置に対して複数のSATAディスク装置を並列に割り当てるにより、アクセス速度の差違を補償し、スペアリング時におけるストレージ装置1000の性能低下を抑制することができる。また、SATAディスク装置はFCディスク装置に比べて信頼性が劣ることがある。従って、FCディスク装置をSATAディスク装置でスペアリングする場合には、FCディスク装置の同じデータを複数のSATAディスク装置にコピーしてもよい。即ち、FCディスク装置をSATAディスク装置でスペアリングする場合には、スペア用のSATAディスク装置を他のSATAディスク装置でミラーリングしてもよい。

【0060】

一方、SATAディスク装置に障害が生じている場合には（ステップS52）、コントローラ310は、複数のFCディスク装置を直列に割り当てることでスペアリングを実行する（ステップS56）。図中に、このスペアリングの様子を模式的に示した。SATAディスク装置がRAIDを構成し、FCディスク装置がスペアとして残っているとする。この状態で、SATAディスク装置のひとつに障害が生じた場合、コントローラは2つのFCディスク装置を直列的に割り当てる。直列に割り当てるとは、1つのFCディスク装置の容量が一杯になった後、2台目のFCディスク装置を使用することを意味する。SATAディスク装置に対して3台以上のFCディスク装置を割り当ててもよい。

【0061】

一般に、ディスク容量は、FCディスク装置の方がSATAディスク装置よりも小さい。従って、一台のSATAディスク装置に対して複数のFCディスク装置を直列に割り当てるにより、容量の差違を補償し、スペアリング時におけるストレージ装置1000の性能低下を抑制することができる。

【0062】

コントローラ310は、以上の手順で異種間のスペアリングを実行すると、その結果に応じて、障害管理テーブルを更新し（ステップS58）、異種スペアリング処理を完了する。この処理では、障害が見いだされたディスク装置は、「閉塞」状態とされる。

D3. 障害通知処理：

図9は障害通知処理のフローチャートである。図7のステップS60に相当する処理であり、コントローラ310が障害の発生状況に応じて、管理装置10への障害通知のタイ

ミングを制御する処理である。

【0063】

この処理では、コントローラ310は、まず、「閉塞」状態となっているディスク装置の有無を判断する（ステップS61）。閉塞状態のディスク装置が存在する場合には、コントローラ310は、即時に障害通知を行う（ステップS67）。閉塞状態は、先に説明した通り、異種ディスク装置間でスペアリングを行った場合に対応する。かかるスペアリングでは、図8に示したように複数台のスペアを割り当てたとしても、必ずしも十分にディスク装置間の性能の差違を補償できるとは限らない。従って、コントローラ310から直ちに障害通知を行い、保守を促すことにより、ストレージ装置1000の性能低下を極力回避可能となる。

【0064】

閉塞状態のディスク装置が存在しない場合（ステップS61）、コントローラ310は、次に「疑似閉塞」状態となっているディスク装置の有無を判断する（ステップS62）。かかるディスク装置が存在しない場合には、コントローラ310は障害通知の必要は無いと判断し、この処理を完了する。

【0065】

疑似閉塞のディスク装置が存在する場合には（ステップS62）、一定の条件が満足されるまで障害通知を保留する。先に説明した通り、疑似閉塞は、同種のディスク装置間でスペアリングを行った場合に対応する。かかるスペアリングでは、通常、ストレージ装置1000の性能は保証されるため、障害通知を保留しても実質的な支障は生じない。本実施例では、かかる状況での障害通知を保留することにより、保守に要する負荷軽減を図っている。

【0066】

疑似閉塞時においても、コントローラ310は、他に通知すべき障害が存在する場合（ステップS63）には、それと併せて障害通知を行う（ステップS67）。また、予め決められた定期的な通知タイミングに至った場合（ステップS64）も障害通知を行う（ステップS67）。更に、疑似閉塞の台数が所定値Th1台を超えた時（ステップS65）、およびスペアの残り台数が所定値Th2を下回った時（ステップS66）のいずれかが満たされる時も、障害通知を行う（ステップS67）。これらの条件を考慮することで、疑似閉塞が生じた場合の障害通知が過度に遅れることを回避することができる。

【0067】

以上で説明した実施例のストレージ装置1000によれば、異種のディスク装置間でスペアリングを許容することにより、スペアを有効活用することができる。この結果、スペアの不足による稼働停止を回避することができる。また、障害が発生したディスク装置の状態を、閉塞、疑似閉塞に区分し、この区分に応じて、障害通知の発行タイミングを制御することによって、ストレージ装置1000の性能低下を回避しつつ、保守の負荷を抑制することができる。なお、異なる種類のディスク装置でスペアリングされている場合には、ユーザまたは保守員がディスク装置を交換または追加するときに、閉塞したディスク装置と同一種類のディスク装置で再びスペアリングするようにしてもよい。例えば、RAIDグループがFCディスク装置で構成され、そのFCディスク装置の一部が故障してSATAディスク装置でスペアリングされている場合に、ユーザまたは保守員がその故障（閉塞）したFCディスク装置を取り替える、またはFCディスク装置を追加した場合に、SATAディスク装置を再び取り替えられたFCディスク装置でスペアリングする。この動作は、FCディスク装置を取り替えまたはFCディスク装置の追加をストレージ装置が認識した後に自動的に行うようにしてもよいし、手動で行っても良い。更に、異なる種類のディスク装置でスペアリングされている状態がある場合は、その状態をディスプレイまたはディスク装置筐体外部等から認識できるようにすることができるよう望ましい。

【0068】

以上、本発明の種々の実施例について説明したが、本発明はこれらの実施例に限定されず、その趣旨を逸脱しない範囲で種々の構成を探ることができることはいうまでもない。

例えば、SATAディスク装置をFC-A Lに接続するための回路、図4に示したDPA 232やインターフェース接続装置233、234は、ディスク装置筐体220側に備えてもよい。実施例では、閉塞状態については、即時に障害通知を行う場合を例示したが（図9のステップ61）、このタイミングは「即時」である必要はなく、疑似閉塞時の通知タイミングよりも短い範囲で任意に設定可能である。

【図面の簡単な説明】

【0069】

- 【図1】実施例としての情報処理システムの概略構成を示す説明図である。
- 【図2】ディスク装置筐体200の斜視図である。
- 【図3】ディスク装置筐体200の内部構造を模式的に示す説明図である。
- 【図4】ストレージ装置1000の内部構造を模式的に示す説明図である。
- 【図5】ディスク種別管理処理のフローチャートである。
- 【図6】障害管理テーブルの構成例を示す説明図である。
- 【図7】スペアリング処理のフローチャートである。
- 【図8】異種スペアリング処理のフローチャートである。
- 【図9】障害通知処理のフローチャートである。

【符号の説明】

【0070】

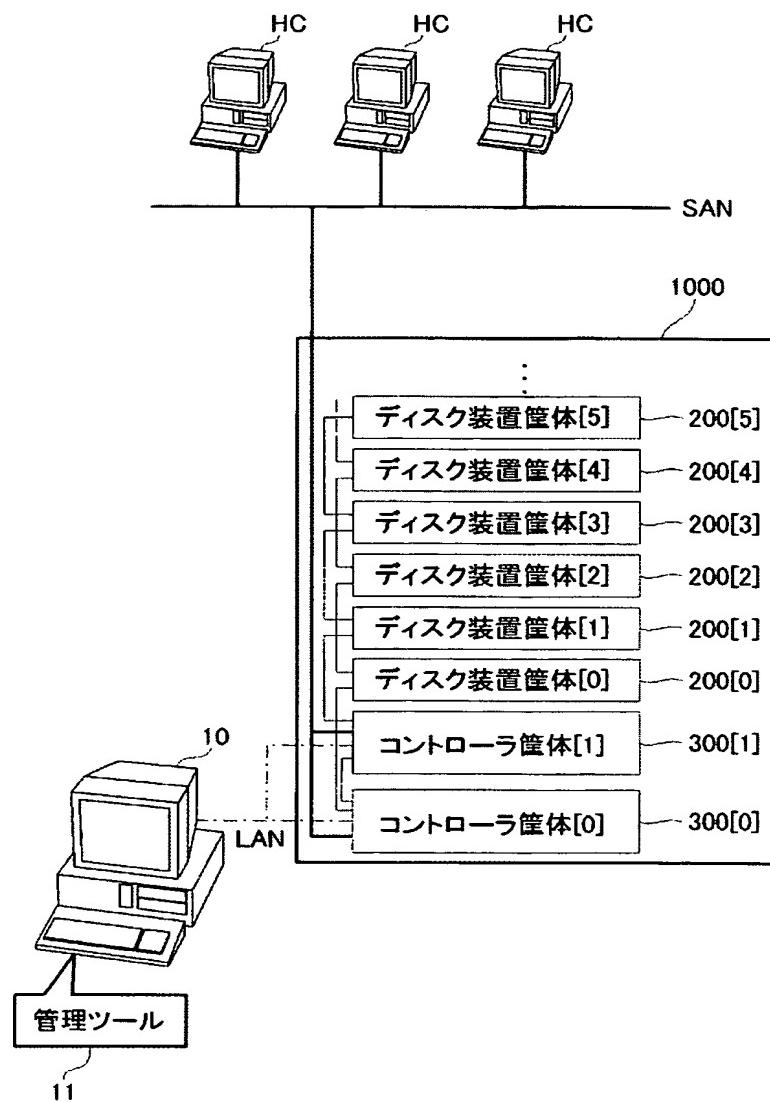
- 10...管理装置
- 11...管理ツール
- 200、200A、200B...ディスク装置筐体
- 210...ルーバ
- 220...ディスク装置
- 220F...FCディスク装置
- 220S...SATAディスク装置
- 221F、221S、231F、231S...コネクタ
- 222F、222S...ハンドル
- 230...バックボード
- 232...DPA
- 233、234...インターフェース接続装置
- 241、242...筐体管理部
- 251、252...PBC
- 300...コントローラ筐体
- 310...コントローラ
- 313...キャッシュメモリ
- 400...搭載ユニット
- 401...本体
- 402、403...ガイド
- 410...基板
- 411、421、413...テーパピン
- 415、416...コネクタ
- 420、420F、420S...ディスク装置
- 421...コネクタ
- 430...キャリー
- 430a...突起
- 430c...コネクタ孔
- 430s...スリット
- 431、432...レール
- 431a、431b...ネジ穴
- 433...ハンドル

434...ラッチ
436、437、438...孔
439...蓋
440...板
441...端部
450、450A...アダプタ
451、451A...本体
452、452A...基板
453、453A、454、454A...コネクタ
455...サブ基板
456、457...コネクタ
458...スペーサ
459...ネジ
1000...ストレージ装置

【書類名】 図面

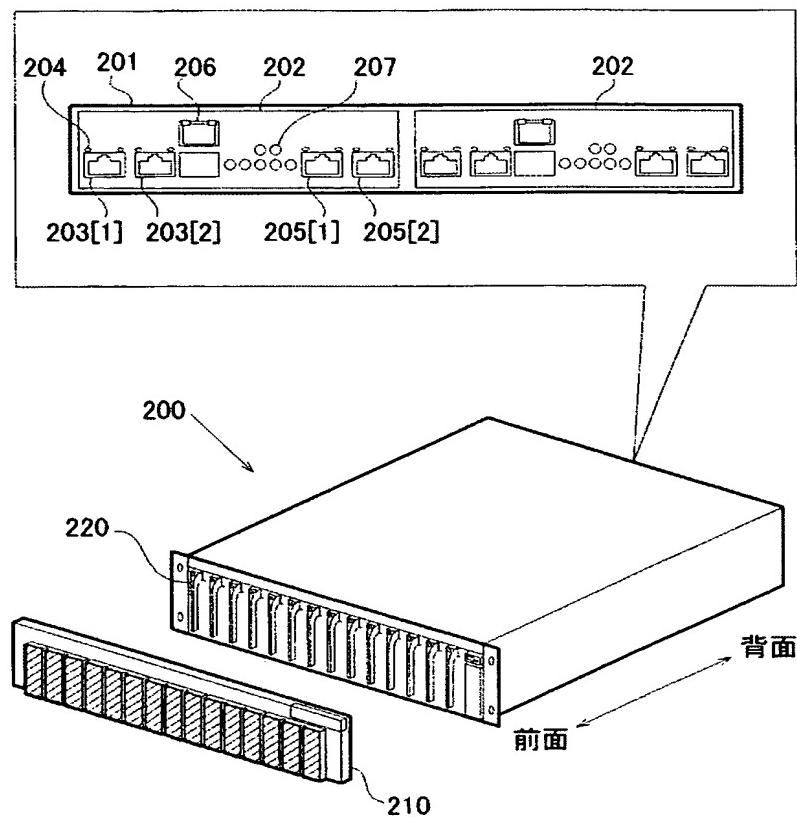
【図 1】

【図 1】



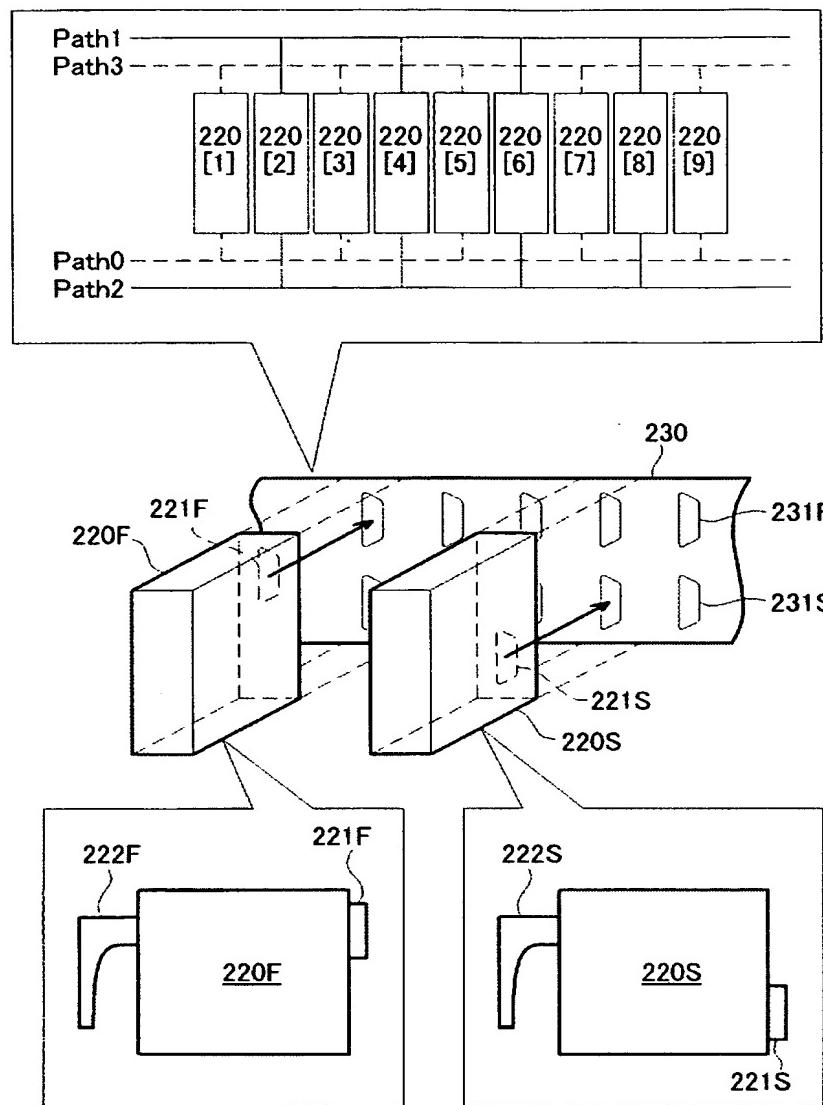
【図2】

【図2】



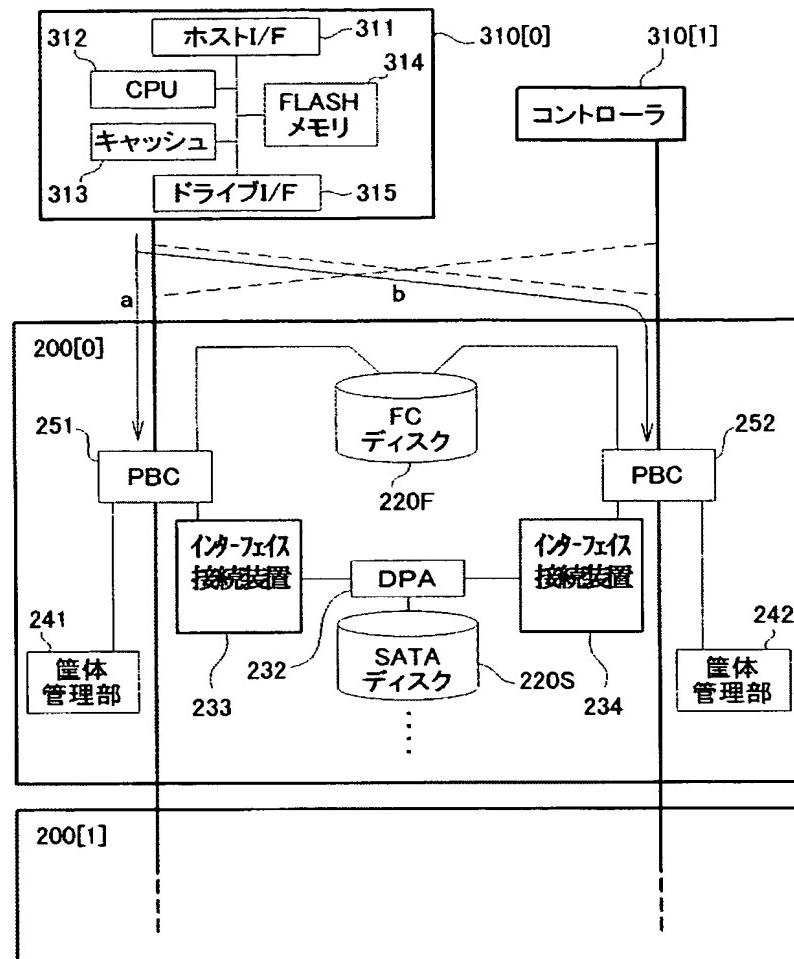
【図3】

【図3】



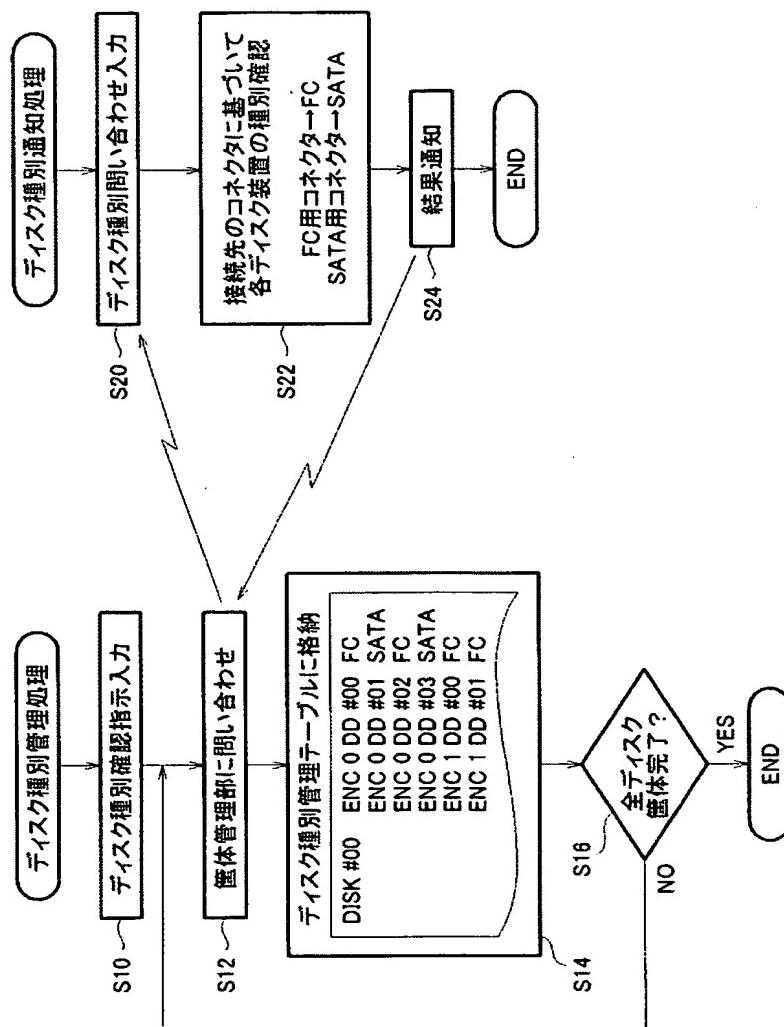
【図4】

【図4】



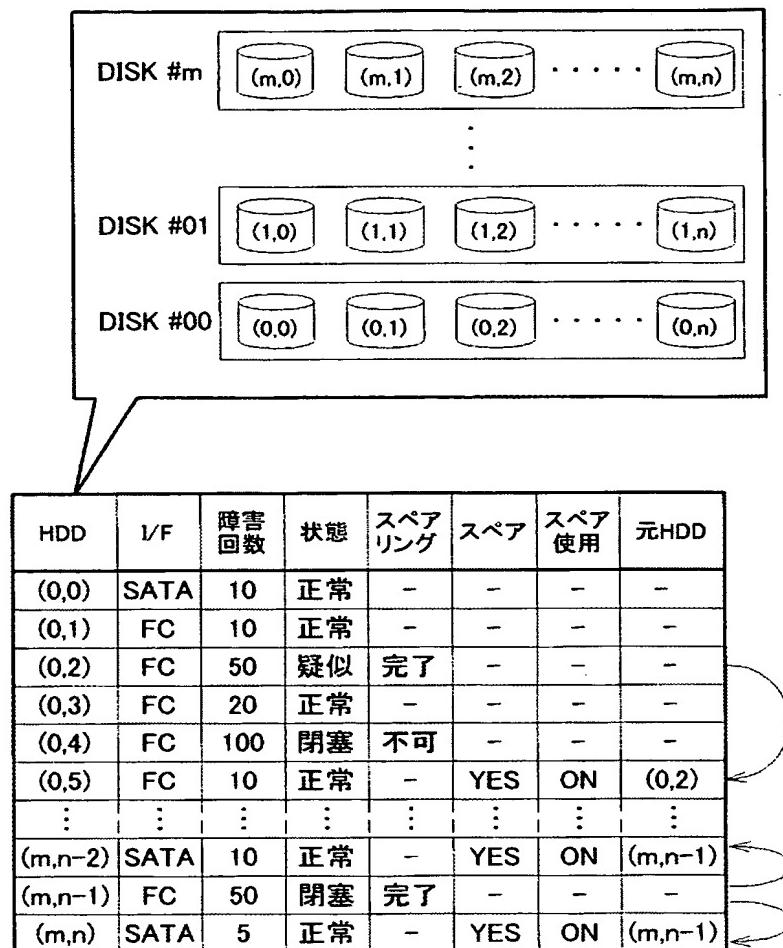
【図 5】

【図 5】



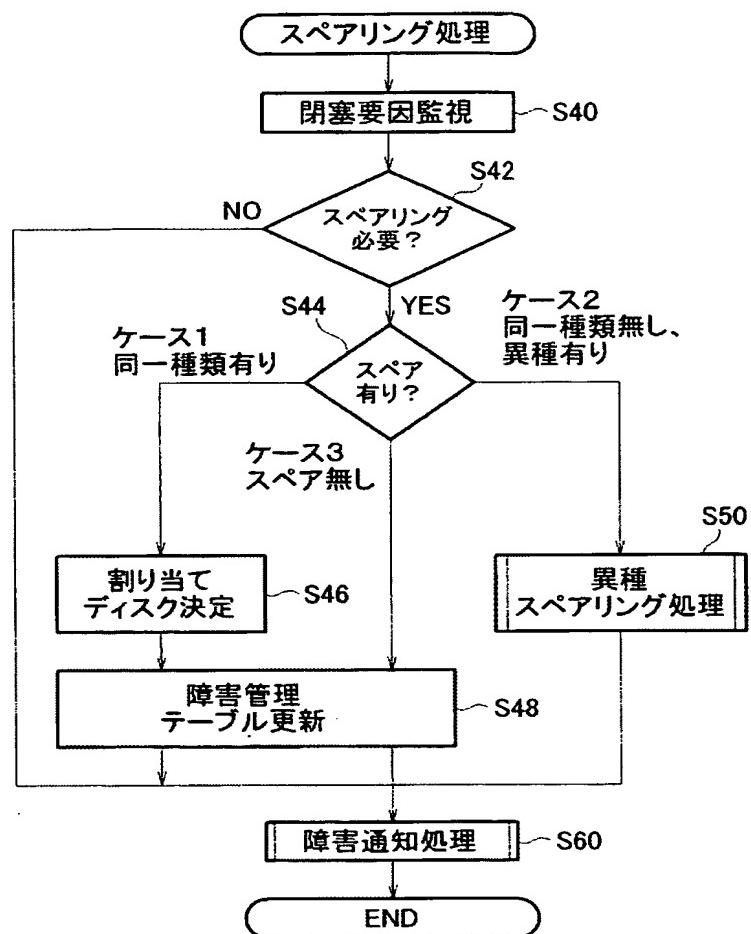
【図6】

【図6】



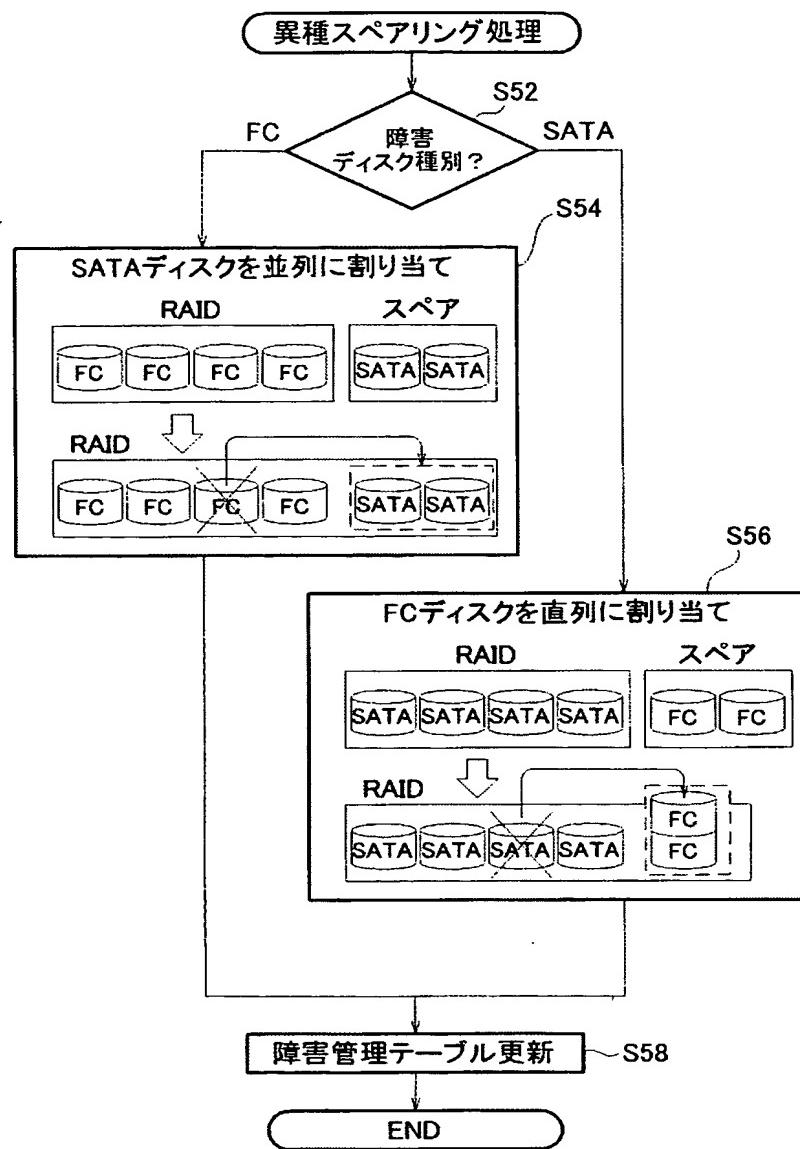
【図 7】

【図 7】



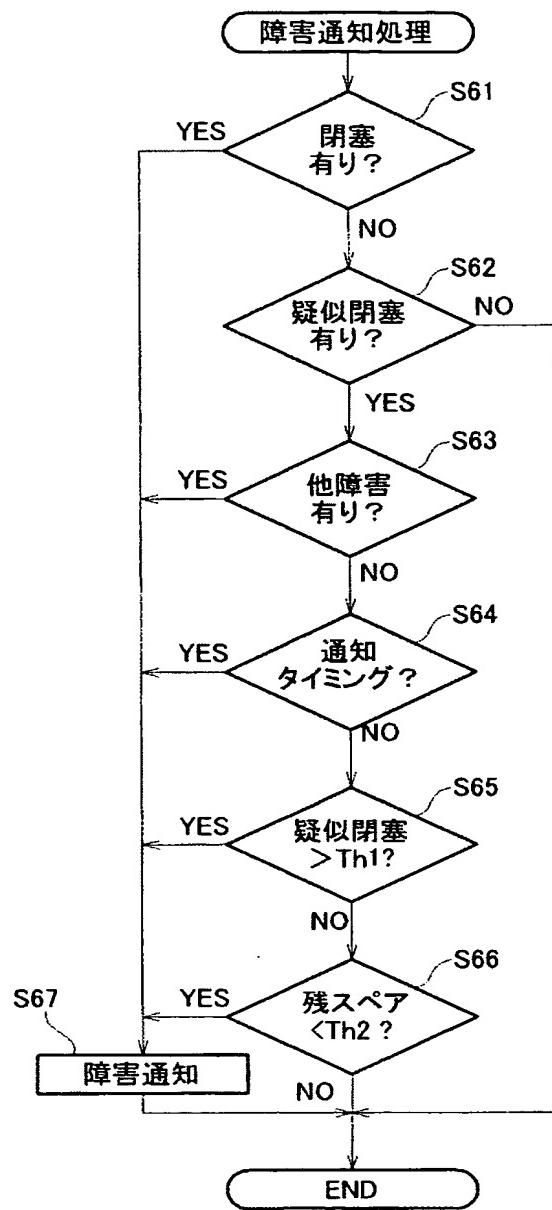
【図8】

【図8】



【図9】

【図9】



【書類名】要約書

【要約】

【課題】 インタフェースが異なる複数種類のディスク装置を混在して格納するストレージ装置において、効率的にスペアリングを実行する。

【解決手段】 ストレージ装置筐体に、ファイバチャネルインターフェースを有するF Cディスク装置220F、シリアルインターフェースを有するS A T Aディスク装置220Sを混在して格納する。コントローラは、アクセス時におけるエラーが所定数を超えたディスク装置について、予め用意されたスペアへのスペアリングを行う。異種のディスク装置間でスペアリングが行われた場合には、即時に障害通知を発信する。同種のディスク装置間でスペアリングが行われた場合には、障害通知を保留する。このように異種間のスペアリングを行うことでスペアの有効活用を図ることができ、障害通知のタイミングを制御することで保守の負荷軽減を図ることができる。

【選択図】 図4

認定・付加情報

特許出願の番号	特願2004-027490
受付番号	50400178389
書類名	特許願
担当官	第七担当上席 0096
作成日	平成16年 2月 5日

<認定情報・付加情報>

【提出日】	平成16年 2月 4日
-------	-------------

特願 2004-027490

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住所 東京都千代田区神田駿河台4丁目6番地
氏名 株式会社日立製作所